

## Research papers

# Uncertainty quantification and partition for multivariate risk inferences through a factorial multimodel Bayesian copula (FMBC) system

Y. Fan<sup>a,\*</sup>, L. Yu<sup>b</sup>, X. Shi<sup>c</sup><sup>a</sup> Department of Civil and Environmental Engineering, Brunel University, London, Uxbridge, Middlesex UB8 3PH, United Kingdom<sup>b</sup> School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China<sup>c</sup> School of Interdisciplinary Studies, University of Glasgow, Dumfries DG1 4ZL, UK

## ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Qiang Zhang, Associate Editor

## Keywords:

Hydrological risk  
Multivariate analysis  
Uncertainty partition  
Factorial analysis

## ABSTRACT

In this study, a factorial multimodel Bayesian copula (FMBC) method is proposed to investigate various uncertainties in the copula-based multivariate risk models and further track the major contributors to the imprecise predictions for different risk indices. In FMBC, the copula models with different marginals and dependence structures will be firstly established with the parameter uncertainties being quantified by the adaptive Metropolis algorithm. A multilevel factorial analysis approach is then adopted to characterize the individual and interactive effects of marginals, copula functions and the associated parameter uncertainties on different risk indices. Moreover, a copula-based dependent sampling algorithm is proposed in the factorial analysis process to generate parameter samples under consideration of their correlation. The applicability of the FMBC approach is demonstrated through the multivariate flood risk analysis at two gauging stations in Wei River basin. The results indicate that extensive fluctuation exists in the inferences of multivariate return periods resulting from different marginal and dependence structures as well as the associated parameter uncertainties. For risk indices of the failure probabilities in AND and Kendall, their predictive variability can be mainly attributed to the uncertainties in model parameters and copula structure, with their total contribution more than 75%. In comparison, the failure probability in OR would be mainly influenced by the parameter uncertainty and also marginal structures, with their total contribution more than 80%. The copula structure would not have a visible effect on the failure probability in OR, with its contribution less than 5% for most scenarios. The obtained results can provide scientific support for reliable hydrological risk inferences within a multivariate context.

## 1. Introduction

Floods are among some of the costliest natural disasters, which cannot only result in severe direct damages and fatalities, but also have considerably wider and longer-term adverse economic consequences (e. g. Kidson and Richards, 2005; Fan et al., 2015; Zhuang et al., 2018; Li et al., 2020; Moeini and Soltani-nezhad, 2020). Assessment and management of flood risks are of great importance for many socio-economic activities especially under changing climate conditions. Nevertheless, floods are generally associated with correlated hydrometeorological processes, which also present multiple attributes such as peak discharge, hydrograph volume and duration (Sarhadi et al., 2016). Univariate risk analyses would mainly focus on one specific flood variable (e.g. flood peak) and can hardly exhibit the properties of other variables, and thus they may be unable to derive reliable inferences for flood events.

Consequently, multivariate approaches are desired since they can provide full description for the occurrence of a flood (The European Parliament and The Council, 2007; Chebana and Ouada, 2011; Requena et al., 2013; Salvadori et al., 2016).

The copula method is one of the most popular approaches for multivariate hydrological risk analyses due to its easiness of implementation and capability of describing complex dependence among correlated variables. Since the first application of copulas in hydrologic simulation (De Michele and Salvador, 2003), a great number of studies have been reported for the development and applications of copula methods in hydrology and geosciences. Some examples include multivariate risk analyses for hydroclimatic extremes such as floods and droughts (Kao and Govindaraju, 2010; Sraj et al., 2014), streamflow simulation (Lee and Salas, 2011; Fan et al., 2017), climate downscaling and projection (Zhou et al., 2018; Sun et al., 2019; 2021).

\* Corresponding author.

E-mail address: [yurui.fan@brunel.ac.uk](mailto:yurui.fan@brunel.ac.uk) (Y. Fan).<https://doi.org/10.1016/j.jhydrol.2021.126406>

Received 1 February 2021; Received in revised form 5 April 2021; Accepted 28 April 2021

Available online 4 May 2021

0022-1694/© 2021 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

The presence of uncertainties, embedded in model structures and parameters, is one of the major issues for both univariate and multivariate risk inferences of hydrological extremes. Research studies have been reported for evaluating uncertainties in copula-based multivariate risk (CMR) models (Serinaldi, 2013; Dung et al., 2015; Fan et al., 2018), which demonstrated existence of extensive uncertainties in CMR models. However, one major issue to be addressed is to characterize the major sources that lead to uncertainties in multivariate risk inferences. In a CMR model, uncertainties in both model parameters and model structures would lead to varied risk inferences for a specific flood event. Recently, some studies have been proposed to reveal how uncertainties would influence the risk inferences in CMR models. For instance, Guo et al. (2020) proposed a Bayesian information-theoretical approach to disclose design flood uncertainty arising from parameter uncertainty in copula-based design flood estimation. Fan et al. (2020a); (2020b); characterized contributions of parameter uncertainties to the resulting risk inferences from the copula-based multivariate risk models. Nevertheless, most of these studies merely addressed the impacts of parameter uncertainties on the resulting risk predictions. In addition to parameter uncertainties, the model choices for both marginal distributions and dependence structures would also challenge the risk predictions from CMR models. For one specific flood event, different marginals and dependence structures may lead to different risk inferences. Moreover, the compound effects of uncertainties in model choices and parameters may dramatically intensify the ambiguity of resulting risk inferences

from CMR models. Therefore, it is desired to address the compound uncertainties in model choices and parameters and further reveal which one would dominate the risk uncertainties from CMR models

Consequently, this study aims to propose a factorial multimodel Bayesian copula (FMBC) system to quantify uncertainties in multivariate risk inferences and further partition the dominant sources for these uncertainties. The FMBC approach integrates copula models, Bayesian parameter estimation and multilevel factorial analysis into a framework. In detail, the multivariate risks in terms of joint return periods and failure probabilities are established based on CMR models with different marginals and dependence structures. For each CMR model with specific marginal and dependence structures, their parameter uncertainties are quantified by a Bayesian-based Markov Chain Monte Carlo (MCMC) algorithm. Finally, the contributions of uncertainties in marginals, dependence structures, and model parameters to the resulting risk inferences are quantified by a multilevel factorial analysis method. Flood data at two hydrological gauge stations in China are used to illustrate the applicability of the proposed FMBC system.

## 2. Methodology

The proposed factorial multimodel Bayesian copula (FMBC) system mainly consists of three components, including: (i) establishment of copula-based multivariate risk (CMR) models with different marginal and dependence structures; (ii) quantification of parameter

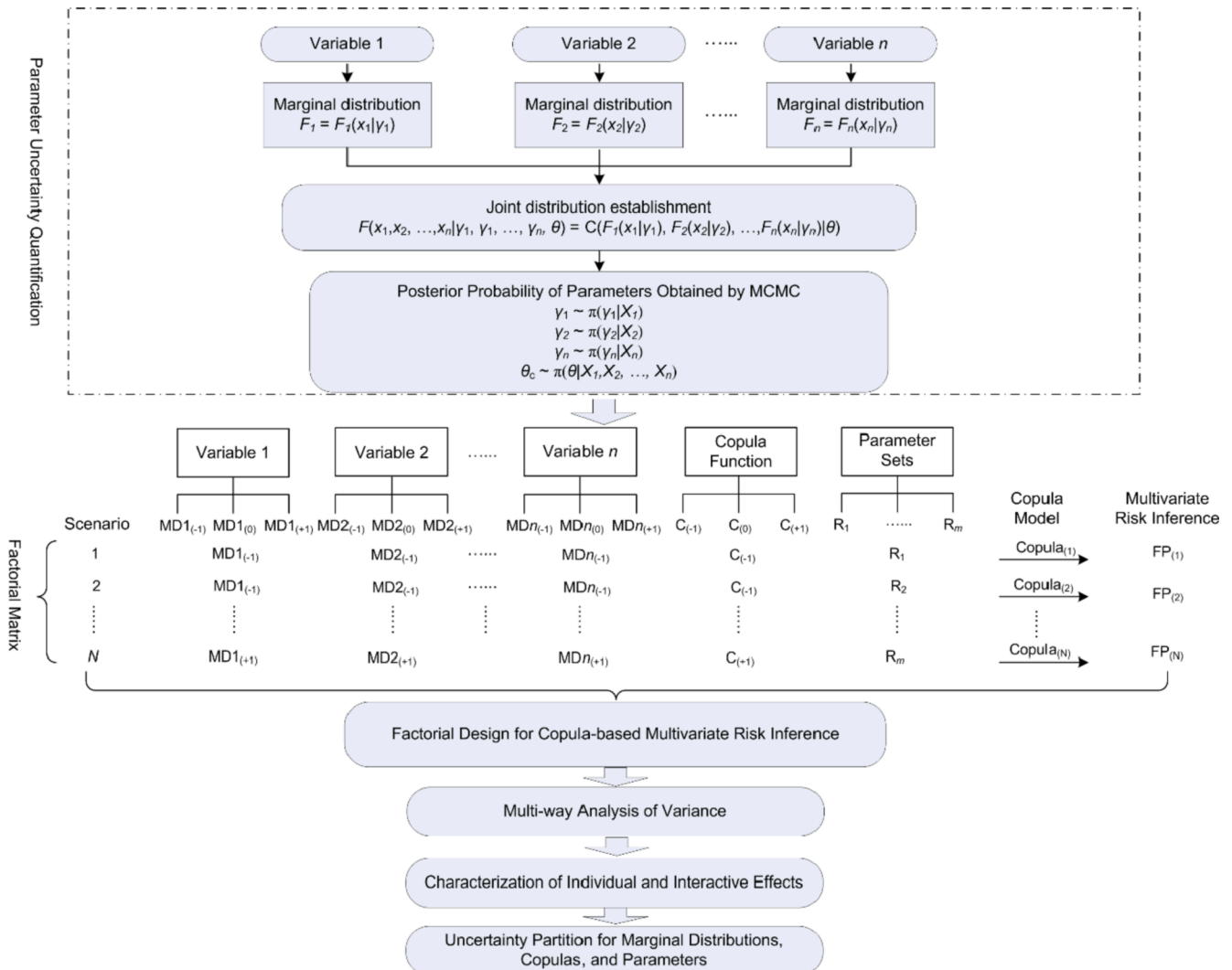


Fig. 1. Framework of the proposed factorial multimodel Bayesian copula approach.

uncertainties for each CRM model; and (iii) characterization of contributions of marginals, copulas and parameters to uncertainties in risk inferences. Fig. 1 illustrates the detailed procedures for FMBC.

### 2.1. Copula-based multivariate risk model

Multiple attributes are generally involved in many hazardous events related to water and hydrological cycles. Typical examples include floods characterized by their peak, volume and duration values (e.g. Xu et al., 2016), and droughts having attributes of severity and duration (e.g. Sun et al., 2019). Moreover, those attributes in a flood or a drought are generally interrelated with each other, leading to a demand for inferring the associated risk in the multivariate context. Copulas, which is initially introduced into hydrology by Favre et al. (2004), have been widely applied for modelling dependent hydrological variables and inferring the risks of compound hydrological events (e.g. Zscheischler and Senéviratne, 2017; Moftakhari et al., 2017). In this part, a general introduction of copula and the associated multivariate risk indices will be provided following a number of research works (e.g. Nelsen, 2006; Salvadori et al., 2007, 2011, 2013, 2016; Graler et al., 2013; Srjaj et al., 2014; Sarhadi et al., 2016; Huang et al., 2017).

For a  $n$ -dimensional random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  with  $F_i(x_i|\gamma_i)$  ( $i = 1, \dots, n$ ) being marginal distributions and  $\gamma_i$  ( $i = 1, \dots, n$ ) being the associated unknown parameters, a copula function can be adopted to build the joint probability distribution for  $\mathbf{X}$  as:

$$F(x_1, \dots, x_n|\gamma_1, \dots, \gamma_n, \theta) = C(F_1(x_1|\gamma_1), \dots, F_n(x_n|\gamma_n)|\theta) \quad (1)$$

where  $C(\cdot)$  is a copula function;  $\theta$  is the parameter in the copula function describing dependence among elements (e.g.  $X_i$ ) of  $\mathbf{X}$ .

In the context of hydrology, one of the most important tasks in risk assessment is to estimate the recurrence interval of one specific hydrologic event (e.g. flood, drought), which is denoted as the return period (RP). However, many extreme events are generally characterized by multiple dependent variables such as peak and volume for a flood, and duration and severity for a drought. This will raise a demand for deriving RP in a multivariate domain to provide robust decision support for water resources management. With the aid of copula function expressed by Eq. (1), three categories of multivariate RP can be derived in terms of “AND”, “OR”, and “Kendall”. Consider a bivariate case which involves two correlated variables. The multivariate RP, respectively denoted as  $T^{OR}$ ,  $T^{AND}$ , and  $T^{Kendall}$  can be formulated as (Salvadori et al., 2011; Fan et al., 2020a; 2020b):

$$T^{OR} = \frac{\mu}{1 - C(u_1, u_2|\theta)} \quad (2)$$

$$T^{AND} = \frac{\mu}{1 - u_1 - u_2 + C(u_1, u_2|\theta)} \quad (3)$$

$$T^{Kendall} = \frac{\mu}{1 - P(C(u_1, u_2|\theta) \leq t)} \quad (4)$$

where  $u_1 = F_1(x_1|\gamma_1)$ ,  $u_2 = F_2(x_2|\gamma_2)$ .

In addition to return period, the risk of failure is characterized as the probability to observe at least one critical event in the design life of one hydraulic infrastructure (Salvadori et al., 2016). Also, based on the copula notation, the concept of failure probability (FP) can be derived in a multivariate context, which are characterized as FP in “AND”, “OR”, and “Kendall”. Consider a critical threshold  $x^* = \{x_1^*, x_2^*\}$  for a bivariate case, the failure probabilities in “AND”, “OR”, and “Kendall” (denoted as  $p_T^{AND}$ ,  $p_T^{OR}$  and  $p_T^{Kendall}$  respectively) can be expressed as (Salvadori et al., 2016):

$$p_T^{OR} = 1 - (C(u_1^*, u_2^*|\theta))^T \quad (5)$$

$$p_T^{AND} = 1 - (u_1^* + u_2^* - \widehat{C}(u_1^*, u_2^*|\theta))^T \quad (6)$$

$$p_T^{Kendall} = 1 - (P(C(u_1^*, u_2^*|\theta) \leq t))^T \quad (7)$$

where  $u_1^* = F_1(x_1^*|\gamma_1)$ ,  $u_2^* = F_2(x_2^*|\gamma_2)$ ,  $(x_1^*, x_2^*)$  defines the bivariate critical threshold.  $T$  indicates the service time of the hydraulic facility under consideration.

Moreover, another common approach to characterize the hazard scenario for an extreme event is to analyze the system under the most likely compound event among feasible combinations with equal return periods (Sadegh et al., 2018). The essence of the method is to select the multivariate hazard event with largest joint probability density which lies on critical layers  $L_q^F$  for a critical level  $q$  (Guo et al., 2020):

$$x^q = \underset{(x_1, x_2 \in L_q^F)}{\operatorname{argmax}} h_C(u_1, u_2|\theta) \quad (8)$$

where  $h(\cdot)$  indicates the density of the copula-based joint probability. The critical layer  $L_q^F$  is defined as:

$$L_q^F = \{(x, y) : C(u_1, u_2|\theta) = q\} \quad (9)$$

where  $u_1 = F_1(x_1|\gamma_1)$ ,  $u_2 = F_2(x_2|\gamma_2)$ .

### 2.2. Uncertainties in the Copula-based multivariate risk models

For both univariate and multivariate hydrologic risk analysis, probabilistic models are desired to quantify the occurrence frequency of flood events and further derive the magnitude of a flood with a specific return period. However, many parametric or non-parametric methods have been proposed for characterizing the probabilistic features of floods in literatures (e.g. Karmakar and Simonovic, 2009). Parametric distributions for flood frequency analysis mainly include extreme value type-1 (EV1), extreme value type 2 (EV2), generalized extreme value (GEV), two-parameter lognormal (LN), Pearson Type III (P3), and log-Pearson Type III (LP3) (Cunnane, 1989; Vogel and Wilson, 1996; Read and Vogel, 2015). Also, some non-parametric techniques, such as kernel density estimation, Gaussian mixture model, and entropy-based approaches are also proposed for flood risk analysis (e.g. Sun et al., 2019). Qi et al. (2016) summarized the three main approaches for distribution selection including (i) official recommendation, (ii) experience knowledge-based selection and (iii) statistical test-based selection. That study also concluded that many candidate distributions can be considered as applicable models, leading to uncertainty in probability distribution selection. In this study, the GEV, P3 and LN distributions will be employed to quantify the probability features of individual flood variables, with their expressions and parameters listed in Table 1. These three distributions are chosen in this study for two reasons: i) based on our previous work (Fan et al., 2020b), the GEV and LN distributions would generally have better performances than the other models, and thus these two distributions would be adopted. ii) the P3 distribution is recommended by the governmental regulation (Ministry of Water Resources of the People's Republic of China, 2006) and thus this distribution will also be selected.

In addition to the uncertainty in marginal distribution selection, there are also many candidate copulas to describe the dependence among correlated flood variables. Some copula functions can be classified as into different families such as elliptical and Archimedean families with each one having different copulas (Brechmann and Schepsmeier, 2013). For instance, a bivariate Archimedean copula can generally be defined as (Nelsen, 2006):

$$C_\theta(u_1, u_2) = \phi^{-1}(\phi(u_1) + \phi(u_2)) \quad (10)$$

where  $u_1$  and  $u_2$  is a specific value of  $U_1$  and  $U_2$ , respectively;  $U_1 = F_{X_1}(x_1)$  and  $U_2 = F_{X_2}(x_2)$ ;  $F_{X_1}$  and  $F_{X_2}$  is the cumulative distribution function (CDF) of random variable  $X_1$  and  $X_2$ , respectively;  $\phi$  is the copula generator that is a convex decreasing function with  $\phi(1) = 0$  and  $\phi^{-1}(t) = 0$  when  $t \geq \phi(0)$ ; the subscript  $\theta$  of copula  $C$  is the parameter hidden in

**Table 1**  
Univariate distributions for individual flood variables.

Name	Probability density function	Parameters	location: $\mu$	scale: $\sigma$
GEV	$\frac{1}{\sigma} \exp\left(-\left(1 + k \frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}}\right) \frac{1}{k} \left(1 + k \frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}-1}$	shape: $k$		
Lognormal	$\frac{1}{x\sqrt{2\pi\sigma_y}} \exp\left(-\frac{(\ln x - \mu_y)^2}{2\sigma_y^2}\right)$ $y = \log(x)$ , $x$ greater than 0, $-\infty < \mu_y < \infty$ , $\sigma_y$ greater than 0	meanlog $\mu_y$		sdlog: $\sigma_y$
Pearson Type III	$\frac{1}{b^a \Gamma(a)} (x-a)^{a-1} e^{-\frac{x-a}{b}}, \Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$	shape: $a$	scale: $b$	location: $a$

the generating function. The Archimedean copulas are quite attractive in multivariate risk analysis, because they can be easily generated and are capable of capturing wide range of dependence structures with several desirable properties, such as, symmetry and associativity (Ganguli and Reddy, 2013). There are also a number of copulas in the Archimedean family with different expressions. In this study, the Gumbel, Frank and Clayton copulas will be adopted, with their properties listed in Table 2.

### 2.3. Bayesian inference of parameter estimation

Once the marginal and copula functions are specified in a CMR model, the parameter uncertainties in both marginals and copula functions are required to be well quantified. In this study, the parameter uncertainty in a CMR model will be quantified by Bayesian analysis. The Bayesian approach has been widely applied for model inference and uncertainty quantification in different fields including multivariate hydrologic risk inferences (Fan et al., 2018; 2020a). Given the prior probability of a certain hypothesis and observations, the posterior distribution can be derived through Bayes' Theorem expressed as:

$$\pi(\theta|\tilde{\mathbf{Y}}) = \frac{L(\theta|\tilde{\mathbf{Y}})\pi_0(\theta)}{\int L(\theta|\tilde{\mathbf{Y}})\pi_0(\theta)d\theta} \quad (11)$$

where  $\pi_0(\theta)$  signifies the prior parameter distribution, and  $L(\theta|\tilde{\mathbf{Y}})$  denotes the likelihood function,  $\int L(\theta|\tilde{\mathbf{Y}})\pi_0(\theta)d\theta$  is the normalization constant,  $\pi(\theta|\tilde{\mathbf{Y}})$  is the posterior probability density function, and  $\tilde{\mathbf{Y}} = \{\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_m\}$  are the observations. Eq. (11) is usually difficult to solve analytically, and thus numerical methods are adopted to generate

evolution (DE) (ter Braak and Vrugt, 2008) and Differential Evolution Adaptive Metropolis (DREAM) methods (Vrugt et al., 2009). In this study, the AM method is to be adopted to approximate the posterior distributions for the parameters in the marginal and copula functions in a CMR model. This method is one of the most common MCMC methods and has been widely used for uncertainty quantification in hydrological simulation and risk analysis (e.g. Viglione et al., 2013; Liu et al., 2020). Details for the AM method can be found in relevant literatures (e.g. Haario et al., 2001).

### 2.4. Uncertainty partition through multilevel factorial analysis

In the context of flood risk assessment through the copula-based multivariate risk (CMR) models, the inferred risk for a specific flood event would be subject to various uncertain factors such as the marginals and copulas, as well as the associated parameters in the CMR model. Consequently, one critical issue to be address is that, among these uncertain factors, which one would have a dominant contribution to the resulting risk inference. Consequently, a multivariate factorial analysis approach is integrated into the MFBC method to answer the above question.

In a factorial analysis, an experimental design is employed to account for all combinations of the levels of factors to help visualize the single effects and interactive effects of factors on a response variable (Fan et al., 2020a). Consider the CMR model with two correlated variables, the effect model of an experimental design on the predictive risk (denoted as  $R$ ) subject to the marginal distributions and copulas can be expressed as:

$$R_{ijkl} = \mu + m_i^C + m_j^A + m_k^B + (m^C m^A)_{ij} + (m^C m^B)_{ik} + (m^A m^B)_{jk} + (m^C m^A m^B)_{ijk} + \varepsilon_{ijkl} \quad (12)$$

$$\begin{cases} i = 1, 2, \dots, c \\ j = 1, 2, \dots, a \\ k = 1, 2, \dots, b \\ l = 1, 2, \dots, n \end{cases}$$

the posterior distributions. The Markov Chain Monte Carlo (MCMC) techniques have been widely adopted to approximate the posterior distributions for parameters. There are a number of MCMC techniques such as Metropolis-Hastings (MH) (Metropolis et al., 1953; Hastings, 1970), adaptive Metropolis (AM) (Haario et al., 2001), differential

where  $R_{ijkl}$  is the total variability for the risk inferences which can be either the joint RP or FP in AND, OR and Kendall;  $\mu$  is the overall mean effect;  $m_i^C$  is the effect of the  $i$ th level of the copula model  $m^C$ ;  $m_j^A, m_k^B$  respectively indicates the effect of the marginals  $m^A$  and  $m^B$  at their  $j$ th and  $k$ th level;  $(m^C m^A)_{ij}, (m^C m^B)_{ik}$  and  $(m^A m^B)_{jk}$  are interactions between factors  $m^C$  and  $m^A$ ,  $m^C$  and  $m^B$ , as well as  $m^A$  and  $m^B$ , respectively;  $(m^C m^A m^B)_{ijk}$  denotes the interaction of factors  $m^C$ ,  $m^A$  and  $m^B$ ;  $\varepsilon_{ijkl}$  denotes the random error component.

In model (12), all the factors  $m^A, m^B$ , and  $m^C$  are fixed and their levels can be either numerical or non-numerical. In this study, the levels (i.e.  $j = 1, 2, \dots, a$ ;  $k = 1, 2, \dots, b$ ) of the factors  $m^A$  and  $m^B$  indicate different choices for the marginal distributions (i.e. P3, GEV, and LN) in the bivariate CMR model, while the levels (i.e.  $i = 1, 2, \dots, c$ ) of the factor  $m^C$  denote different choices for the copula functions (i.e. Gumbel, Clayton

**Table 2**  
Basic properties of applied copulas.

Copula Name	Function $[C_\theta(u_1, u_2)]$	$\theta \in$	Generating functions $[\phi(t)]$
Clayton	$[u_1^{-\theta} + u_2^{-\theta} - 1]^{-1/\theta}$	$[-1, \infty) \setminus \{0\}$	$\frac{1}{\theta}(t^{-\theta} - 1)$
Gumbel	$\exp\{ -[(-\ln u_1)^\theta + (-\ln u_2)^\theta]^{1/\theta} \}$	$[1, \infty)$	$(-\ln t)^\theta$
Frank	$-\frac{1}{\theta} \ln\{1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1}\}$	$[-\infty, \infty) \setminus \{0\}$	$-\ln[\frac{e^{-\theta t} - 1}{e^{-\theta} - 1}]$



and Frank). Moreover, model parameters would also influence the risk inferences in the CMR model. Nevertheless, model parameters in a CMR model are not independent with the factors for marginals and the copula function. They are associated with the choice of marginals and copulas. Therefore, the parameters can hardly be considered as independent factors in model (12). In order to characterize the effects of parameter uncertainties on the risk inferences, these effects are considered as the random error component in model (12). In detail, for a CMR model (with specific choices for marginals and copula), a number of samples for its parameters are generated from their posterior distributions quantified by AM method. The corresponding risk inferences are then generated from the CMR model based on these parameter samples.

Based on the effect model denoted by Eq. (12), the total variability of the predictive risk can be decomposed into its component parts as follows:

$$SS_T = SS_C + SS_A + SS_B + SS_{AC} + SS_{BC} + SS_{AB} + SS_{ABC} + SS_e \quad (13a)$$

and

$$SS_T = \sum_{i=1}^c \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}^2 - \frac{R^2}{abcn} \quad (13b)$$

$$SS_C = \frac{1}{abn} \sum_{i=1}^c R_{i...}^2 - \frac{R^2}{abcn} \quad (13c)$$

$$SS_A = \frac{1}{bcn} \sum_{j=1}^a R_{j..}^2 - \frac{R^2}{abcn} \quad (13d)$$

$$SS_B = \frac{1}{acn} \sum_{k=1}^b R_{...k}^2 - \frac{R^2}{abcn} \quad (13e)$$

$$SS_{AC} = \frac{1}{bn} \sum_{i=1}^c \sum_{j=1}^a R_{ij..}^2 - \frac{R^2}{abcn} - SS_C - SS_A \quad (13f)$$

$$SS_{BC} = \frac{1}{an} \sum_{i=1}^c \sum_{k=1}^b R_{i.k.}^2 - \frac{R^2}{abcn} - SS_B - SS_C \quad (13g)$$

$$SS_{AB} = \frac{1}{cn} \sum_{j=1}^a \sum_{k=1}^b R_{jk.}^2 - \frac{R^2}{abcn} - SS_A - SS_B \quad (13h)$$

$$SS_{ABC} = \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^a \sum_{k=1}^b R_{ijk.}^2 - \frac{R^2}{abcn} - SS_C - SS_A - SS_B - SS_{AC} - SS_{BC} - SS_{AB} \quad (13i)$$

$$SS_e = \sum_{i=1}^c \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}^2 - \frac{1}{n} \sum_{i=1}^c \sum_{j=1}^a \sum_{k=1}^b R_{ijk.}^2 \quad (13j)$$

where  $R_{ijk.} = \sum_{l=1}^n R_{ijkl}$ ,  $R_{ij..} = \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}$ ,  $R_{j.k.} = \sum_{i=1}^c \sum_{l=1}^n R_{ijkl}$ ,  $R_{i.k.} = \sum_{j=1}^a \sum_{l=1}^n R_{ijkl}$ ,  $R_{i...} = \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}$ ,  $R_{j..} = \sum_{i=1}^c \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}$ ,  $R_{...k} = \sum_{i=1}^c \sum_{j=1}^a \sum_{l=1}^n R_{ijkl}$ ,  $R_{...} = \sum_{i=1}^c \sum_{j=1}^a \sum_{k=1}^b \sum_{l=1}^n R_{ijkl}$ . Then the contributions of marginal distributions, dependence structures and parameter uncertainties on risk inferences can be calculated as:

(1) Contributions of marginal distributions A and B

$$\eta_A = SS_A / SS_T \quad (14a)$$

$$\eta_B = SS_B / SS_T \quad (14b)$$

(2) Contribution of the parameter in the dependence structure

$$\eta_C = SS_C / SS_T \quad (14c)$$

(3) Contribution of model interactions among marginals and copula

$$\eta_{AC} = SS_{AC} / SS_T \quad (14d)$$

$$\eta_{BC} = SS_{BC} / SS_T \quad (14e)$$

$$\eta_{AB} = SS_{AB} / SS_T \quad (14f)$$

$$\eta_{ABC} = SS_{ABC} / SS_T \quad (14g)$$

(4) Contribution of parameter uncertainties

$$\eta_e = SS_e / SS_T \quad (14h)$$

## 2.5. Procedures for the proposed FMBC system

The proposed FMBC system integrates the copula-based multivariate risk inference model, MCMC method, and multilevel factorial analysis to characterize the major and interactive effects of model selection and parameter uncertainties on the resulting risk inferences. The detailed procedures of the FMBC method are as follows:

**Step 1:** For one set record of extreme events, choose the marginals and the dependence structures to establish the CMR models. In this study, the candidate marginals include P3, GEV and LN distributions while the candidate copulas include Gumbel, Clayton and Frank copulas.

**Step 2:** For the CMR model with specific marginals and dependence structure, estimate parameter posteriors through the AM method stated in Section 2.2.

**Step 3:** Establish the factorial matrix consisting of all the combinations for the marginals and copulas. If there are three candidates for marginals and copulas, the factorial matrix will contain 27 combinations for the bivariate CMR model, as presented in Table 3.

**Step 4:** For each CMR model with known marginals and dependence structure (i.e. one row in Table 3), sample the parameter values from their posterior distributions obtained in Step 2. In this step, the parameter values are sampled based on a dependent sampling algorithm since the parameters in the CMR model may be correlated with each other. We argue that overlook for their interdependence may overestimate impacts of parameter uncertainties on the resulting risk inferences from CMR. The dependent sampling algorithm is based on the

**Table 3**

The matrix of all combinations for different marginal and copula functions.

Marginal 1	Marginal 2	Copula
GEV	GEV	Frank
P3	P3	Gumbel
GEV	LN	Clayton
P3	LN	Clayton
LN	GEV	Frank
P3	P3	Frank
LN	P3	Clayton
P3	P3	Clayton
P3	LN	Frank
GEV	LN	Gumbel
GEV	GEV	Gumbel
LN	P3	Frank
LN	LN	Frank
LN	GEV	Gumbel
LN	LN	Gumbel
LN	LN	Clayton
P3	GEV	Frank
P3	GEV	Gumbel
GEV	P3	Gumbel
GEV	LN	Frank
LN	GEV	Clayton
P3	GEV	Clayton
GEV	GEV	Clayton
P3	LN	Gumbel
GEV	P3	Frank
LN	P3	Gumbel
GEV	P3	Clayton

multivariate copula theory, in which the dependence among model parameters are quantify through a vine copula model. Such a sampling algorithm can be performed as follows:

4a). For the parameter vector  $\Theta = (\gamma^A, \gamma^B, \theta^C)$ , where  $\gamma^A, \gamma^B$  represent the parameters in the marginal distributions and  $\theta^C$  indicates the parameter in the copula model, quantify their marginal posterior distributions through nonparametric kernel estimator. For instance, the posterior distribution for  $\gamma^A$  can be denoted as  $\gamma^A \sim G(\gamma)$ . The original parameter values can be converted into uniform values in  $[0, 1]$  (denoted as  $u$ ) through their posterior distributions.

4b). Establish the joint probability distribution for the parameter vector through a vine copula model as follows:

$$c_{1:D}(\mathbf{u}|\eta) = \prod_{j=1}^{D-1} \prod_{i=1}^{D-j} c_{i,i+j|i+1,\dots,i+j-1}(F(u_i|u_{i+1}, \dots, u_{i+j-1}), F(u_{i+j}|u_{i+1}, \dots, u_{i+j-1})|\eta) \quad (15)$$

where  $D$  is the total parameters in the parameter vector  $\Theta$ .  $\eta$  is the parameter vector for the vine copula model  $\mathbf{u} = (u_1, u_2, \dots, u_D) \in [0, 1]^D$

4c). Produce the sample set  $\mathbf{u}^i = (u_1^i, \dots, u_D^i) (i = 1, 2, \dots, ns)$  through simulation of the copula model denoted in Eq. (15). There are several algorithms to generate samples from a multivariate copula model (Aas et al., 2009). Such a process can also be implemented based on the R package “CDVine” (Brechmann and Schepsmeier, 2013).

4d). Generate the new samples for the original parameters through their inverse probability function  $G^{-1}(u)$ .

Step 5: Based on sampled parameters for the CMR model obtained in Step 4, the risk inferences, in terms of both joint RP and FP in AND, OR and Kendall can be generated based on Eqs. (2)–(7), which can be expressed in a generic form as:

$$R = H(x_1^*, x_2^* | \Theta^i)$$

where  $H$  is a CMR model with specific marginal and copula structures and  $\Theta^i (i = 1, 2, \dots, ns)$  are the associated parameter values samples in Step 4.

Step 6: Repeat Steps 4 and 5 for all the marginal and copula combinations in the factorial matrix (i.e. Table 3).

Step 7: Calculate the total variability (i.e.  $SS_T$ ) of the risk inferences and its decomposition components expressed in Eqs. (13).

Step 8: Generate both the individual and interactive effects of marginals, dependence structure and the associated parameters on the multivariate risk inferences through Eqs. (14).

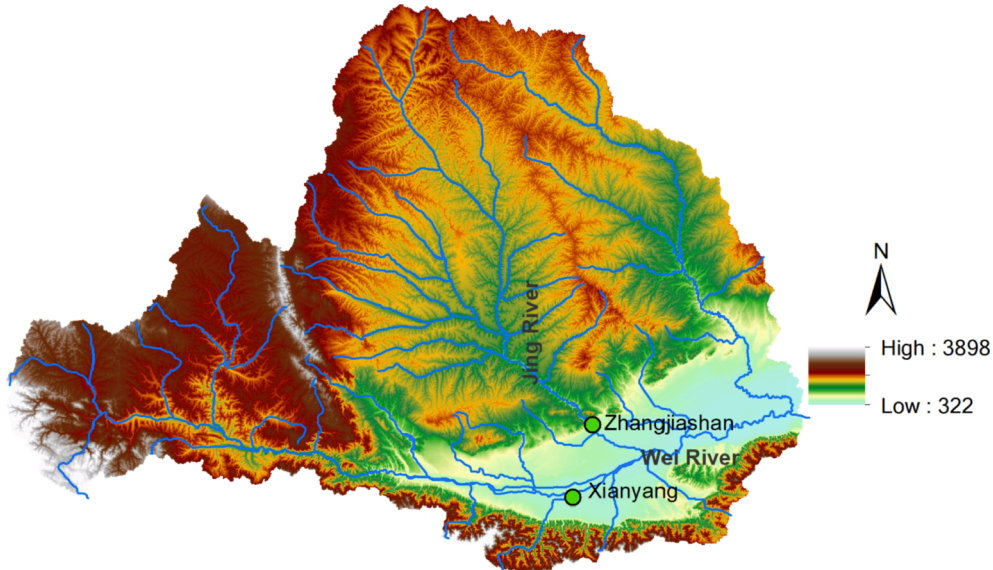
### 3. Applications

The applicability of the FMBC approach will be demonstrated at two streamflow gauging stations in the Wei River basin. The Wei River is the largest tributary of the Yellow River, originating from the Niaoshu Mountain in the Weiyuan County of Gansu province, having a main-stream length of 818 km and a drainage area of nearly 135,000 km<sup>2</sup> (Song et al., 2007; Zuo et al., 2014; Du et al., 2015; Xu et al., 2016). The regional climate in this area is semi-arid and sub-humid continental monsoon, with significant temporal-spatial variations in precipitation (Xu et al., 2016). The annual mean precipitation is about 559 mm but shows a noticeable decrease from south to north, with the annual precipitation fluctuating between 800 and 1000 mm in the south part, and between 400 and 700 mm in the north part (Xu et al., 2016). The mean temperature in the entire Wei River basin varies from 6 to 14 °C, the annual potential evapotranspiration fluctuates from 660 to 1,600 mm, and the annual actual evapotranspiration is about 500 mm (Du et al., 2015).

Streamflow data observed at the Xianyang and Zhangjiashan stations will be adopted for the demonstration of the proposed FMBC method. As presented in Fig. 2, the Xianyang station is located at the mainstream of the Wei River and the Zhangjiashan station is located at the Jing River, which is one of the major tributaries of Wei River. Based on the observations, annual maximum data series will be employed in this study, with the associated flood volumes are obtained through the method

**Table 4**  
Flood characteristics for different stations.

Station name	period		flood variable	
			Peak (m <sup>3</sup> /s)	Volume (m <sup>3</sup> /(s day))
Xianyang	1960–2006	Minimum	139	317
		Median	1350	2491
		Maximum	12,380	17,802
		Mean	1734.6	3592.8
		Std	1819.9	3606.8
		Kurtosis	26.0	6.3
Zhangjiashan	1958–2012	Minimum	217	303.7
		Median	775	1365.3
		Maximum	3730	7576.1
		Mean	904.5	1790.8
		Std	686.0	1609.5
		Kurtosis	6.2	3.1



**Fig. 2.** The location of the studied watershed and the hydrological gauging stations.

proposed by Yue (2000); (2001;). Table 4 presents some descriptive statistics for the flood records at the Xianyang and Zhangjishan stations. There are 47 and 55 flood events characterized at the Xianyang and Zhangjishan stations based on data streamflow observations.

## 4. Results analysis

### 4.1. Parameter estimation for different CMR models

In the hydrological context, we can have multiple choices for both the marginal distributions and dependent structures to formulate the copula based multivariate risk models. As presented in Tables 1 and 2, three distributions including GEV, lognormal (LN) and Pearson Type III distributions are going to be adopted to quantify the random features for individual flood variables (i.e. peak and volume), whilst the Gumbel, Frank and Clayton copulas are employed to describe the dependence between correlate random variables. These candidate marginal and dependence models will lead to a number of CMR models as listed in

Table 3. For each CMR model, its parameters would be another uncertainty source which may lead to imprecise risk inferences and thus need to be well quantified. In this study, the adaptive Metropolis (AM) method has been adopted to quantify the parameter uncertainties in the CMR model with specified marginal and dependence structures.

Fig. 3 presents the posterior distributions for the parameters in the CMR models with different marginal and dependence structures. It is obvious that, even though one parameter (e.g. shape parameter in P3) would have similar posterior distributions, it still shows slightly different probabilistic features in different CMR models. As presented in Fig. 3, the shape parameter in P3 would reach its highest probability value (i.e. PDF) at 2.02 for the CMR model with a structure of P3-P3-Gumbel. However, the highest probability for this parameter is located at about 1.9 and 2.13 for the CMR models with structures of P3-P3-Frank and P3-P3-Clayton. Such a difference is more visible for location parameter in P3 quantifying the distribution of flood volume (i.e. the 6th row in Fig. 3). These differences can also be observed in other CMR models as presented in Fig. S1.

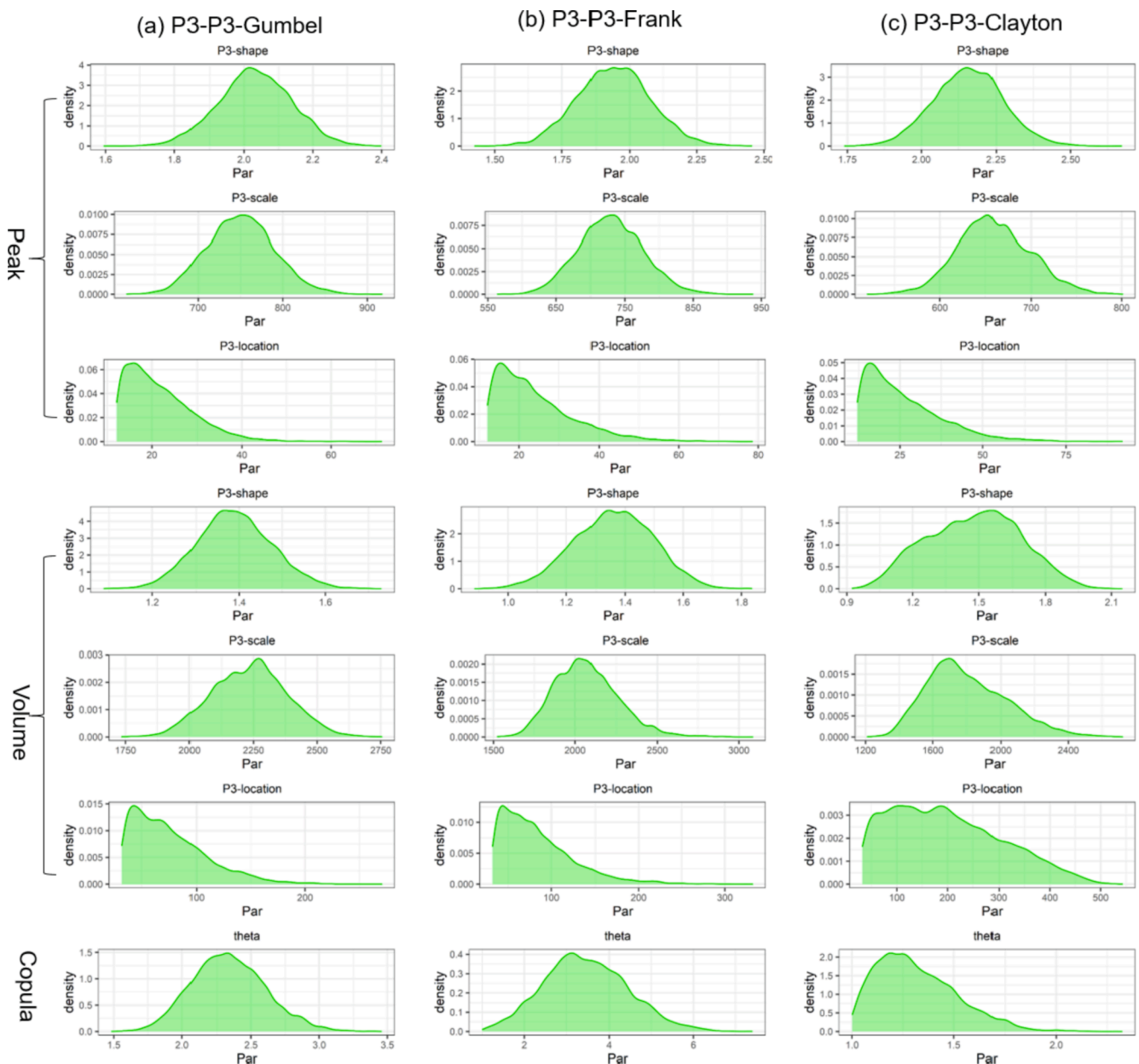


Fig. 3. Posterior distributions of parameters for the CMR models with the same marginals (i.e. P3) but different copulas at Xianyang station.

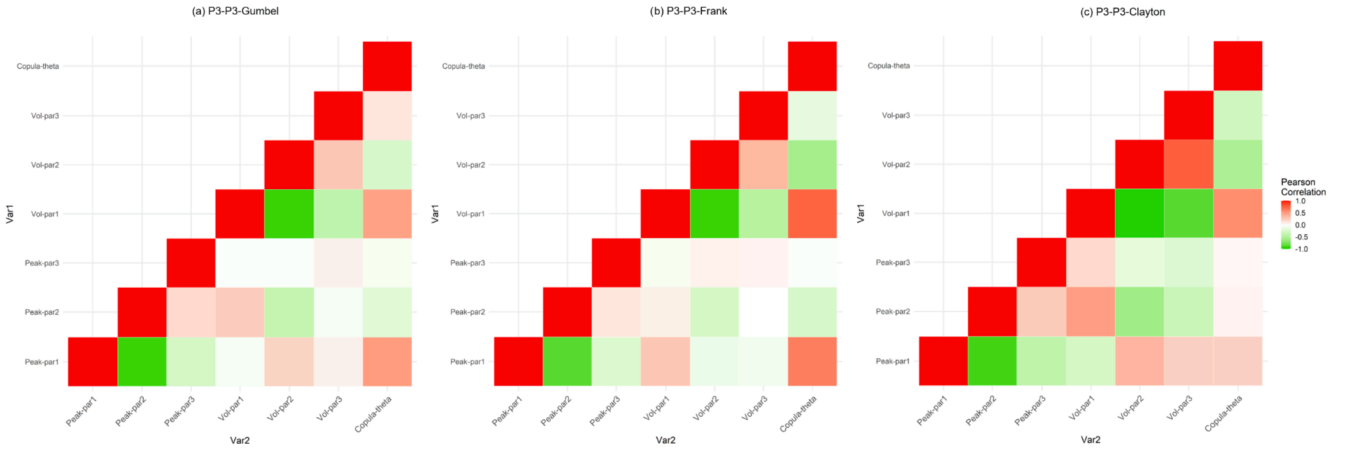


Fig. 4. Parameters' correlation for different CMR models at Xianyang station.

The difference in the posterior distributions for one parameter in different CMR models may result from two causes. Firstly, since the MCMC method is random in nature, the posterior distributions for one parameter can hardly be identical in different runs of the MCMC method. Another reason may be due to the correlation in the parameters in the CMR model. As presented in Fig. 4, some parameters in the CMR model are significantly correlated, which implies that the changes in one parameter may influence the values for others. For different CMR models, they would have different structures. For instance, Fig. 3(a) shows parameter posteriors for the CMR model with a structure of P3-P3-Gumbel while the parameter posteriors in Fig. 3(b) correspond to the CMR model consisting of P3-P3-Frank. The posteriors for the parameter in copula functions are visibly distinguishable due to different function structure between the Gumbel and Frank copula. The difference in the posterior distributions for the parameter in copula functions in these two CMR models (i.e. P3-P3-Gumbel and P3-P3-Frank) would then lead to distinction in parameter posteriors in marginal distributions due to parameters' correlation.

#### 4.2. Uncertainty in risk inferences

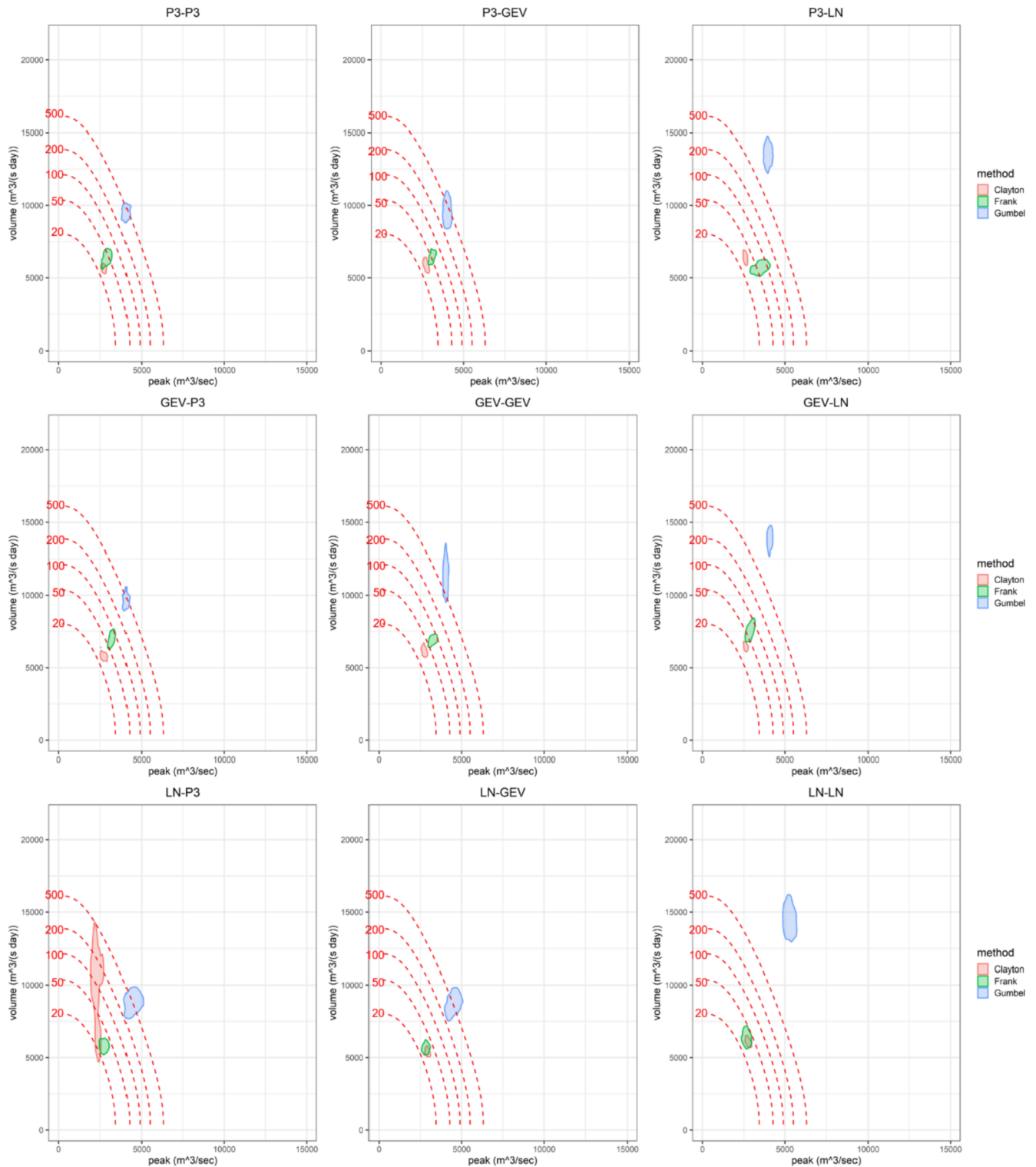
For one specific extreme event, different CMR models may generate different risk inferences due to those uncertainties embedded in marginal and dependence model structures as well as the associated model parameters. Moreover, for different risk indices (e.g.  $T^{OR}$  or  $T^{AND}$  expressed by Eqs. (2) or (3)), the CMR models may produce different predictive uncertainties for them.

Fig. 5 presents the predictive uncertainties for the multivariate RP in AND at Xianyang Station. In this figure, the contour curves with red dashed line represents multivariate RP values averaged from all the CMR models under consideration of different marginal and dependence structures as well as the associated parameter uncertainties. In comparison, the colored contours in each subfigure indicate the most likely compound events with a  $T^{AND}$  of 50-year obtained from the CMR model with different copula functions. In detail, the uncertainties enclosed within the colored contour present the predictions with a confidence level of 80% for the most likely compound events due to the parameter uncertainties in the CMR model. As presented in Fig. 5, the predictions of the most likely flood with a  $T^{AND}$  of 50-year are significantly different from CMR models with different marginals, copula functions and parameter sets. For instance, for the same risk level (e.g.  $T^{AND} = 50$ ), the CMR model having the marginals being P3 and P3 (i.e. Fig. 5(a)) would infer the most likely floods with the average  $T^{AND}$  less than 50 years

when the Clayton copula is adopted, while in comparison it will produce the most likely floods with the average  $T^{AND}$  even larger than 200 years when the Gumbel copula is employed. Even for the CMR model with specified marginal and dependence structures, it may produce imprecise flood predictions due to uncertainties in its parameters as presented in the colored contours. Fig. S2 presents the uncertain risk inferences for the multivariate RP in AND from the CMR models with different marginals, copulas and parameter sets at the Zhangjiashan station. Remarkable uncertainties exist in the predictions for the most likely extremes from different CRM models, as being observed at the Xianyang station.

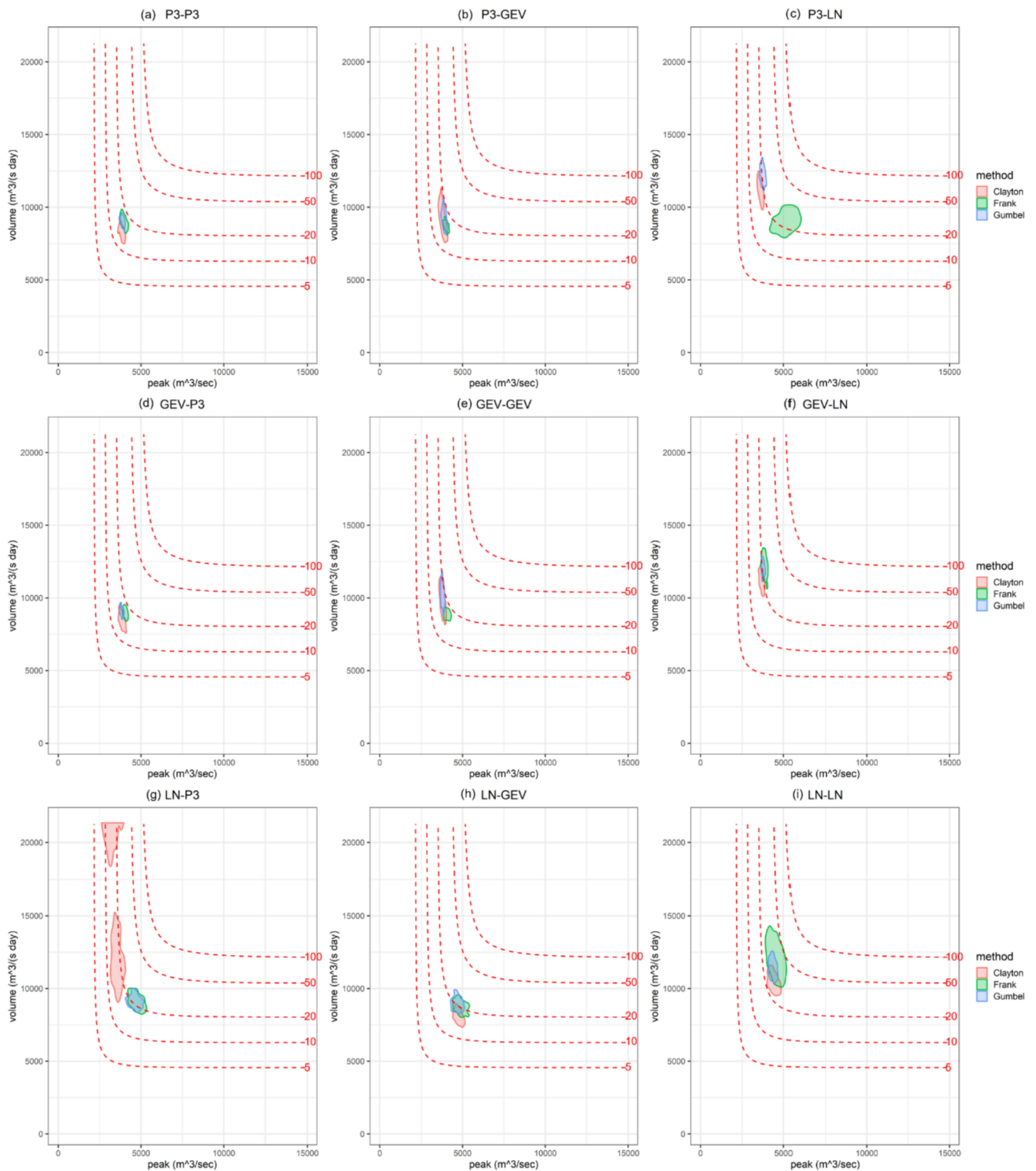
Similar to the multivariate RP in AND, the predictions of multivariate RP in OR for a compound extreme event also presents extensive uncertainties due to the uncertainties embedded in the marginal and dependence structures as well as parameter uncertainties in the CMR models. Fig. 6 shows the contour of the average multivariate RPs in OR from different CMR models, as well as the predictions for the most likely floods with a  $T^{OR}$  of 20-year at the Xianyang station. The results indicate that for one specific risk level (i.e.  $T^{OR} = 20$ ), the predictions for the most likely flood events would be varied remarkably due to the uncertainties in structures and parameters of the CMR model. Moreover, the predictive uncertainties of the most likely floods under the risk index of  $T^{OR}$  show a different pattern from the variations of the predictive floods under  $T^{AND}$ . For instance, consider the CMR model with the two marginals being the P3 distribution, the predictions for the most likely floods under  $T^{AND} = 50$  are highly distinguishable for the CMR models using different copula functions as presented in Fig. 5(a), even though the parameter uncertainties also lead to visible variations in the flood predictions. In comparison, for the most likely floods under  $T^{OR} = 20$ , most predictions are located in the region with the average  $T^{OR}$  ranging between 10 and 20 years, as presented in Fig. 6(a). This implies that for the CMR models with the marginals being the P3 distribution, the selection of the copula function would have less impact on the inferences of multivariate RP in OR than it has on the multivariate RP in AND. Also, similar patterns can be observed for the CMR models with GEV or LN marginals from Figs. 5 and 6.

Figs. S2 and S3 present the predictions of the most likely floods with  $T^{AND}$  and  $T^{OR}$  respectively being 50 and 20 years from the CMR models consisting of different marginals, copula functions and parameter sets at the Zhangjiashan station. The results indicate that the parameters may have significant impacts on the flood predictions for both  $T^{AND}$  and  $T^{OR}$ , which leads to large colored contours in Figs. S2 and S3. Nevertheless, the predictions of the most likely floods under consideration of  $T^{AND}$



**Fig. 5.** Predictive uncertainties for the most likely flood with a 50-year joint return period in AND at Xianyang station. The red dashed contour exhibits the joint return period in AND averaged from all the CMR models. The colored regions represent the most likely flood events resulted from parameter uncertainties of the CMR models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)





**Fig. 6.** Predictive uncertainties for the most likely flood with a 20-year joint return period in OR at Xianyang station. The red dashed contour exhibits the joint return period in OR averaged from all the CMR models. The colored regions represent the most likely flood events resulted from parameter uncertainties of the CMR models. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

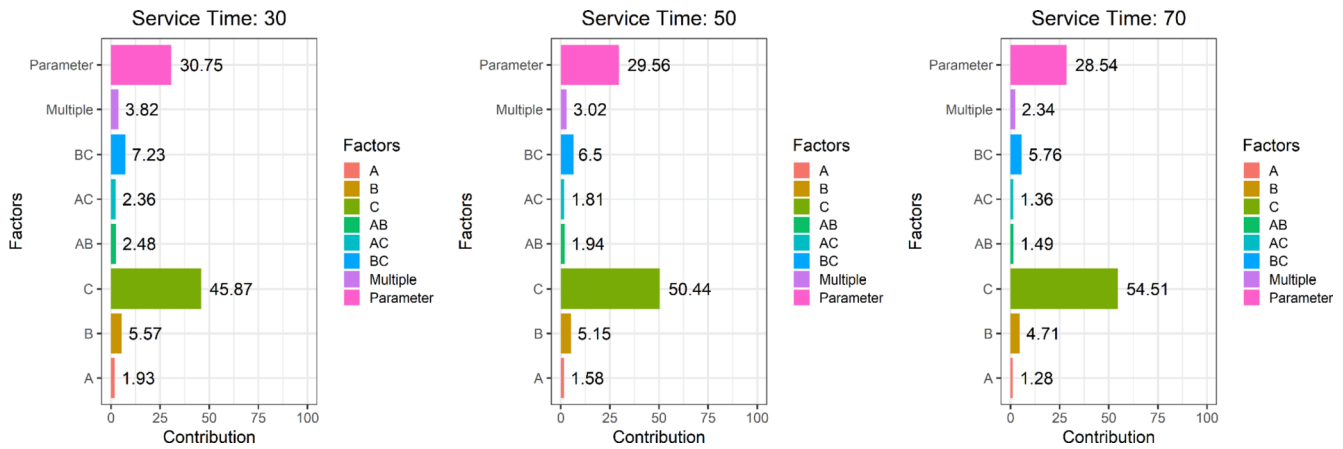


Fig. 7. Contributions of different uncertain factors to the variability of failure probability in AND at the Xianyang station.

from the CMR models with different copulas are still more distinguishable than the flood predictions under  $T^{OR}$ . This also implies a higher contribution of copula selection in the CMR model to the multivariate risk in AND than that risk in OR.

#### 4.3. Contribution partition for uncertainty sources

In the copula-based multivariate risk (CMR) models, various uncertainties in model structures and also parameters may lead to imprecise risk predictions. Moreover, for one specific risk index (e.g. multivariate RP in AND), different uncertainty sources would generally have different impacts on the resulting predictions. Also, for the same uncertainty source, it may have different contributions to the predictions of different risk indices, as presented in Figs. 5 and 6. Consequently, some critical issues to be investigated are, for uncertainties in the marginal and dependent structures as well as the associated model parameters, (i) which one will contribute most for the resulting risk inferences, (ii) whether the contributions will be varied for different risk indices and also the service time of the corresponding hydraulic facilities. To address the above questions, the multilevel factorial analysis (MFA) method is introduced into the developed FMBC system to characterize both the individual and interactive effects of those uncertainty sources on the multivariate risk inferences.

For the factorial analysis in the FMBC system, the two marginals and the dependence model (i.e. copula) are designed as factors to be addressed. In detail, the GEV, P3 and LN distributions are employed to quantify the random features in the individual flood variables, while the Clayton, Gumbel and Frank copulas are adopted to describe the

dependence among the flood variables. This will form a  $3^3$  factorial design consisting of 27 CMR models, as presented in Table 3. Furthermore, the parameter uncertainties, which are quantified by the MCMC method, are associated with the specification of CMR model structure. Thus, the model parameters cannot be considered as one factor in the factorial analysis. Nevertheless, this kind of uncertainties are reflected as the replicates of the factorial experiments. For each CMR model, 10 parameter sets are sampled from their posterior distributions based on the dependent sampling algorithm, which drive the CMR model ten times to generate the relevant risk inferences. The failure probabilities in “AND”, “OR”, and “Kendall” (denoted  $p_T^{AND}$ ,  $p_T^{OR}$  and  $p_T^{Kendall}$  respectively) expressed by Eqs. (5)–(7) are to be derived from the CMR models. Then the contribution for each uncertainty source (i.e. marginals, copulas and parameters) to different failure probabilities (i.e.  $p_T^{AND}$ ,  $p_T^{OR}$  and  $p_T^{Kendall}$ ) can then be characterized by Eqs. (14).

Fig. 7 presents the individual and interactive contributions of various uncertainty sources to the failure probabilities in AND (i.e.  $p_T^{AND}$ ) under different service time scenarios at the Xianyang station. It is visible that different factors have different impacts on the risk index of  $p_T^{AND}$ . In detail, the copula function and the parameter uncertainties would have more impacts on the inference of  $p_T^{AND}$  than other factors, with a total contribution of more than 75%. In comparison, the marginal distributions in the CMR model would not pose significant impacts on the FP in AND, with the highest contribution less than 6%. However, for those two marginals in the CMR model, the marginal selection for flood volume is more influential than that for flood peak. This may due to the fact that the magnitude of the flood volume are generally higher than the corresponding flood peak. Furthermore, the contributions from different

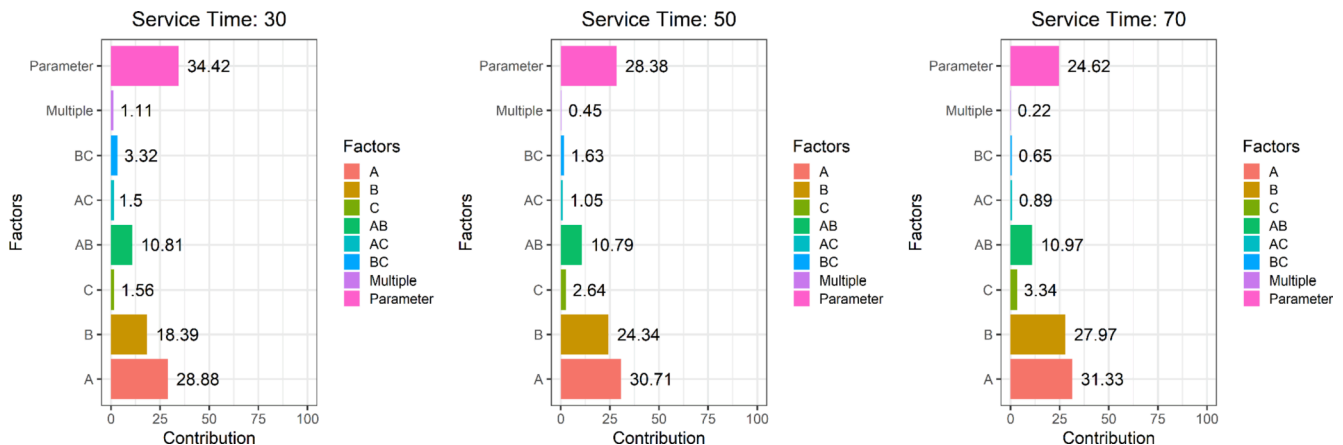


Fig. 8. Contributions of different uncertain factors to the variability of failure probability in OR at the Xianyang station.

factors on  $p_T^{AND}$  also show some variations under different service time scenarios. It is noticeable that the structure of the copula function would have higher contributions with the increase in service time while the contributions from other factors show an opposite trend. For instance, the copula selection will have a contribution of 45.87%, 50.44% and 54.51% to the variability in  $p_T^{AND}$  respectively for a service time of 30, 50 and 70 years, while the contributions from parameter uncertainties are 30.75%, 29.56%, and 28.54, respectively. Fig. S4 presents the contributions of different factors on the variability of the risk index  $p_T^{AND}$  at Zhangjiashan station. Even though the impacts from parameter uncertainties and copula structures have a different rank, all the factors show a similar contribution pattern with their contributions at the Xianyang station. The parameter uncertainties and copula structures would have much higher impacts than the other factors. Also, the contributions from the copula function will increase with the increasing service time while the contributions of parameter uncertainties would decrease simultaneously.

Fig. 8 shows the individual and interactive effects of marginal and dependent structures as well as the parameter uncertainties on the risk index of FP in OR (i.e.  $p_T^{OR}$ ) at Xianyang station. Compared with the FP in AND, the contributions from the studied factors  $top_T^{OR}$  show a different pattern. In addition to the noticeable impact from parameter uncertainties, the marginal distributions also have significant impacts on the risk inference of  $p_T^{OR}$ . Moreover, for the two marginal distributions, the marginal corresponding to the flood peak (i.e. factor A in Fig. 8) would have a higher contribution to the variability of  $p_T^{OR}$  than the marginal distribution for flood volume. This is mainly due to the higher kurtosis of the flood peak as presented in Table 4. This indicates that the flood peak would have greater extremity of outliers and thus the changes of marginal distribution for this variable would lead to noticeable variation in the risk index of  $p_T^{OR}$ . In addition, the contributions of the two marginals will increase as the service time of the hydraulic infrastructures increases while at the same time the impacts from parameter uncertainties will decrease. In terms of the copula function, it will not have visible contributions (less than 4%) to the predictive variability of  $p_T^{OR}$  as it ever did for the risk index of  $p_T^{AND}$  (larger than 45%). Fig. S5 presents the contributions of different factors on the risk inference of  $p_T^{OR}$  at Zhangjiashan Station, which showed a different impact pattern from that at the Xianyang station. The parameter uncertainties would dominate the risk inference of  $p_T^{OR}$  at this station with its contributions more than 80% for different service time scenarios. This would be mainly due to the extensive uncertainties in model

parameters which has also been demonstrated by Fig. S3. Nevertheless, there are still some similar features for the impacts of the studied factors on the risk index of  $p_T^{OR}$  at both Xianyang and Zhangjiashan stations. As the increase of service time, the contribution of parameter uncertainties will decrease and at the same time the contributions of marginals will increase. Also, the marginal distribution for the flood peak would have a higher contribution  $top_T^{OR}$  than the marginal for the other variable.

For the FP in Kendall expressed by Eq. (7), those uncertain factors (i.e. marginal distributions, copulas, and model parameters) have a similar impact pattern with the risk index of  $p_T^{AND}$  as indicated in Fig. 9. The selection of copula function will have the highest contribution to the predictive variability of  $p_T^{Kendall}$ , followed by the parameter uncertainties. Also, the contribution from copula would increase with the increase of service time while at the same time the impact of parameter uncertainties would decrease. However, compared with the FP in AND (i.e.  $p_T^{AND}$ ), the copula function tends to have a higher contribution to the variability of  $p_T^{Kendall}$  with the minimum contribution larger than 50%. Such a conclusion can also be demonstrated by the results at the Zhangjiashan station presented in Fig. S4. Even though more extensive uncertainties have been observed in the model parameters at this station, the contribution from the selection of copula is even higher increasing from 61.2% for a service time of 30 years to 67.1% for a service time of 70 years.

## 5. Discussion

In the proposed FMBC approach, the copula-based dependent sampling algorithm is proposed in order to reflect parameters' interdependence of the CMR model. Negligence of their correlation may overestimate the contributions of parameter uncertainties to the risk inferences. Fig. 10 present the contributions of the uncertain factors (i.e. marginals, copulas and parameters) to the risk inferences for FP in AND, OR and Kendall under the independent sampling scenario, where the model parameters are sampled independently from their posterior distributions. The results indicate that, ignoring parameters' correlation would lead to slightly higher quantification for the impact of parameters' uncertainties on the risk indices of  $p_T^{AND}$  and  $p_T^{Kendall}$ . For instance, the contribution of parameter uncertainties  $top_T^{AND}$  would be about 32.09% for a service time of 30-year under the independent sampling scenarios while, as presented in Fig. 7, such a contribution would be 30.75% when dependent sampling algorithm is adopted. Nevertheless, for the PF in OR (i.e.  $p_T^{OR}$ ), it can be observed that the overlook of parameters' correlation

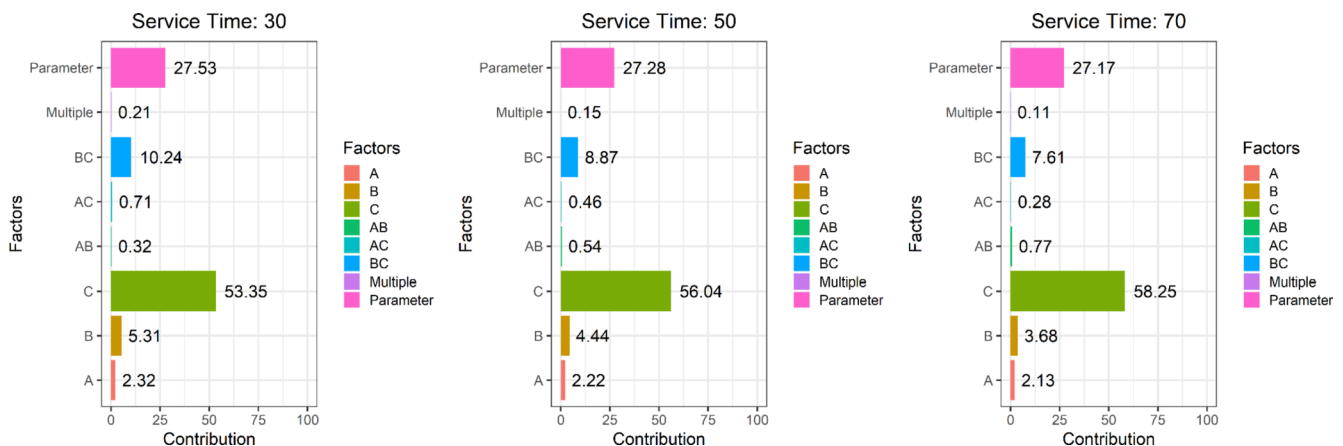


Fig. 9. Contributions of different uncertain factors to the variability of failure probability in Kendall at the Xianyang station.

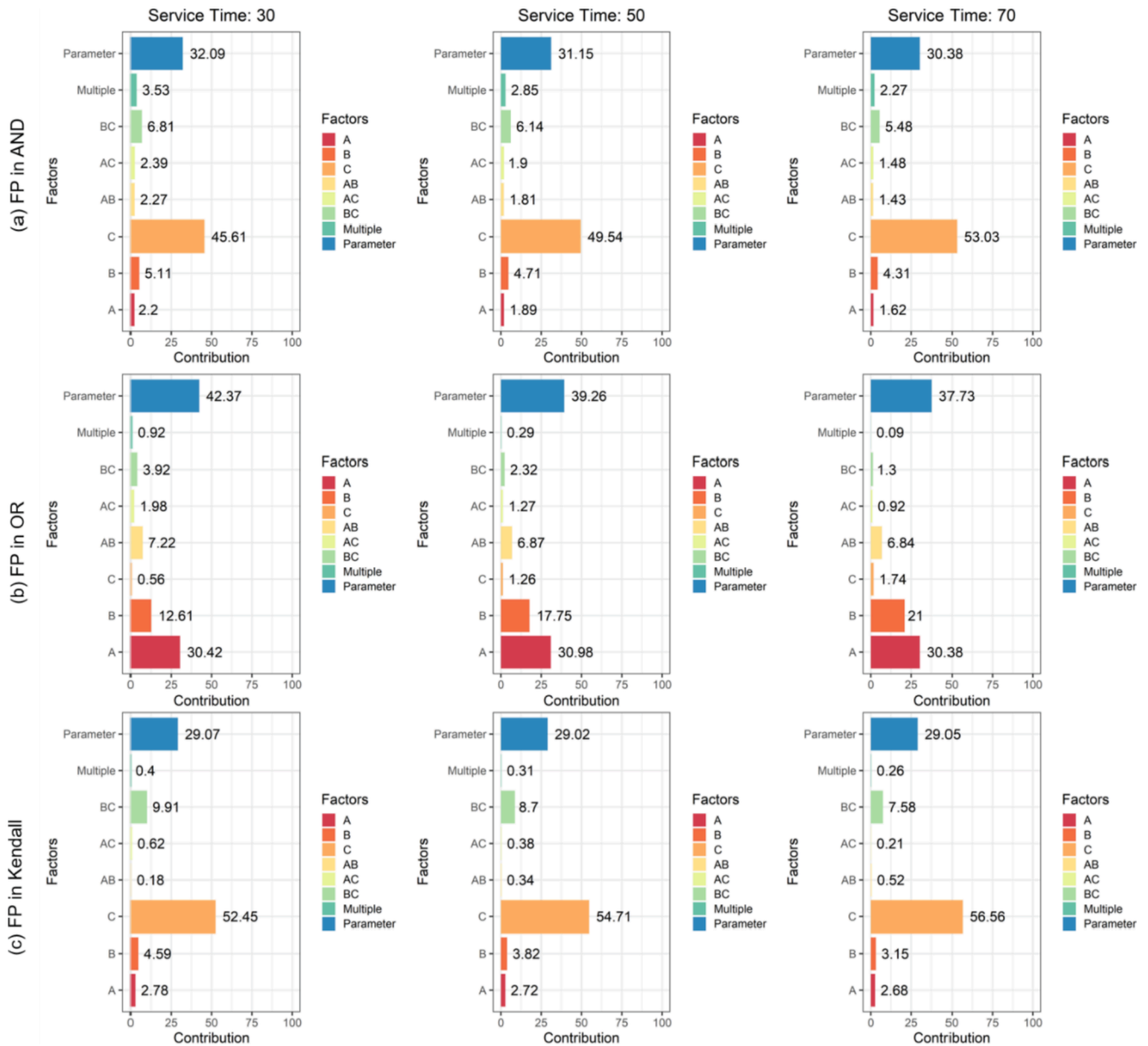


Fig. 10. Contributions of different uncertain factors to the variability of failure probabilities under the independent sampling scenario at Xianyang station.

in the CMR model would noticeably overestimate the impact of parameters' uncertainties on this risk index. As presented in Fig. 10, the contribution of parameters' uncertainty would be 42.4% for a service time of 30-year under the independent sampling scenario while in comparison such a contribution, as shown in Fig. 8, would be about 34.4% for dependent sampling scenarios. Moreover, their difference would be intensified for the increase in the service time. Consequently, the dependent sampling algorithm, developed in the proposed FMBC approach, can provide more reliable results for characterizing contributions of uncertain factors to multivariate risk inferences.

In the proposed FMBC approach, the parameter uncertainties are firstly quantified by AM-based MCMC approach. The copula-based dependent sampling algorithm is proposed to jointly sample the parameter values in the factorial analysis process to reflect impacts of parameter uncertainties. Different sample sizes in this process may lead

to different contributions from model parameters. The results presented in Figs. 7–9 are based on a sample size of 10. Fig. 11 exhibits the contributions of different uncertain factors to the three FP indices based on a sample size of 20. The results suggest that, as the increase of the sample size, there are no evident changes for the contributions of different uncertain factors when compared with the results under the 10-sample scenario. There are slight discrepancies under these two sample size scenarios. For instance, the parameter uncertainty would have a contribution of 30.75%  $top_T^{AND}$  for a service time of 30 years under the sample size of 10, and such a contribution will slightly decrease to 28.65% for the sample size of 20. Nevertheless, these differences would not significantly change the results of contribution partition of different uncertain factors to the multivariate risk inferences.

The proposed FMBC approach cannot only be applied for the multivariate flood risk inferences for the annual maxima of peak

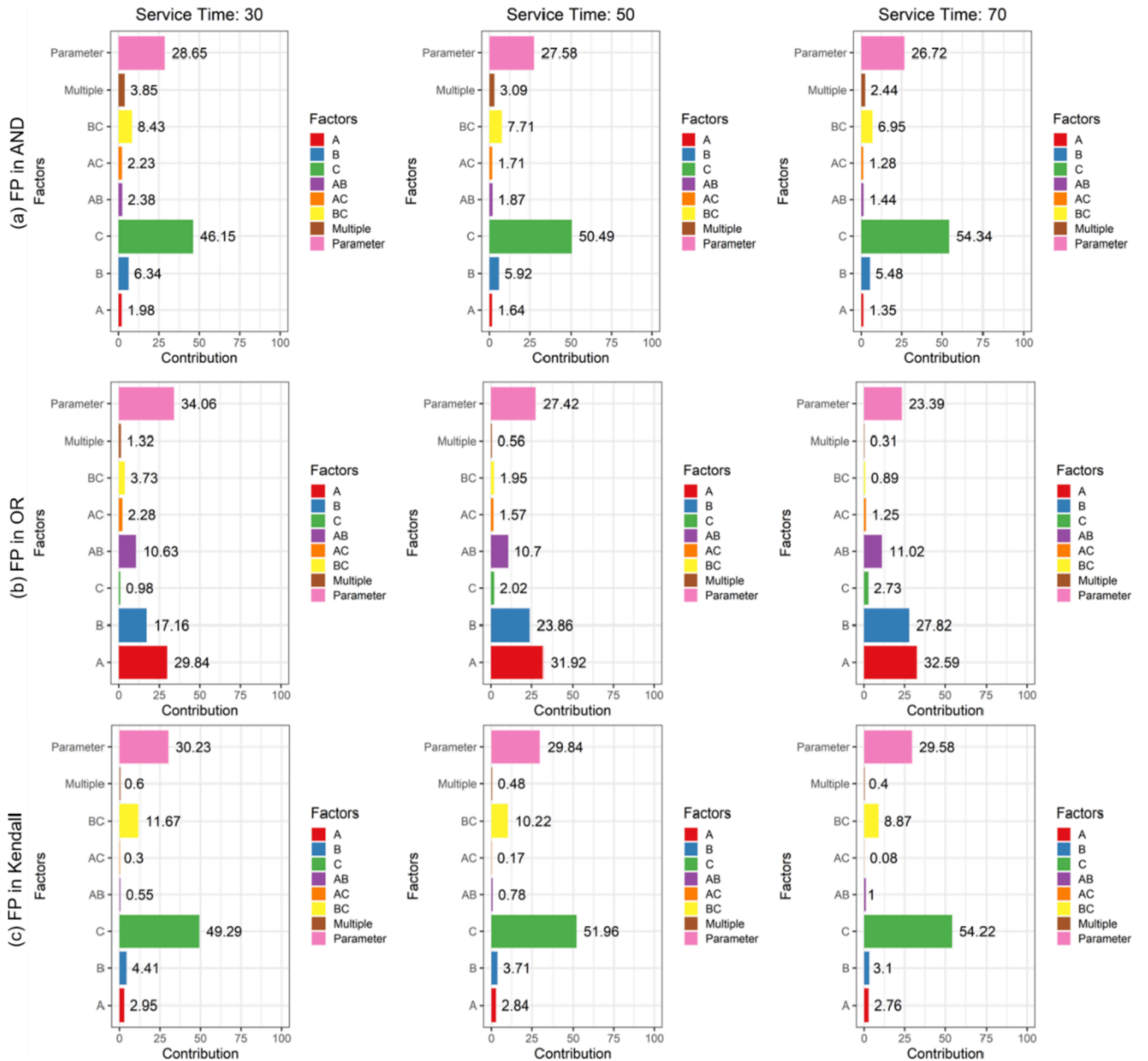


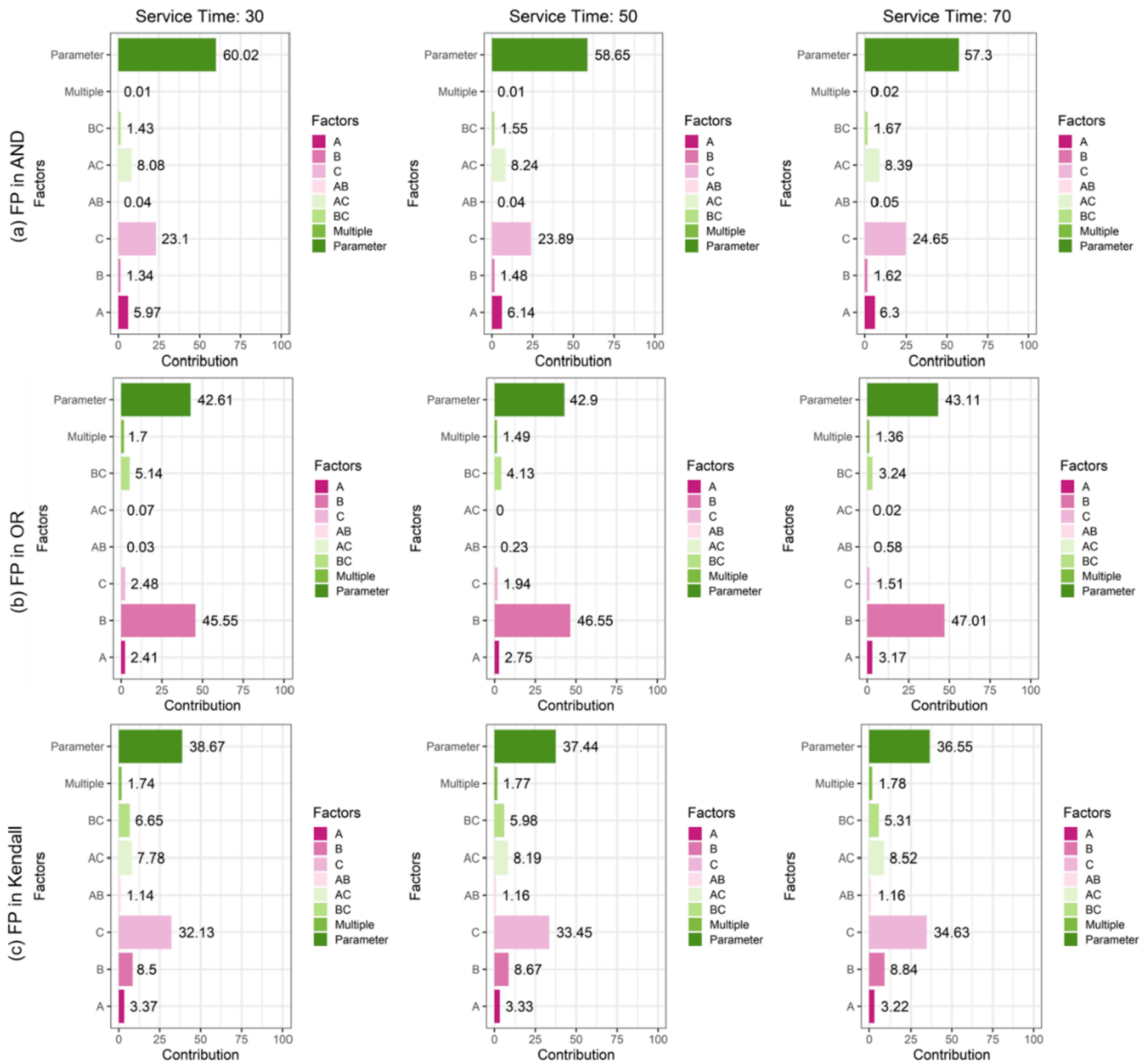
Fig. 11. Contributions of different uncertain factors to the variability of failure probabilities under a sample size of 20 at Xianyang station.

discharge associated with its flood volume, but also be applicable for other flood variables. Fig. 12 present the contributions of model structures (i.e. marginal and dependence structure) and parameter uncertainties to the three FPs for the annual maxima of 3-day volumes and their associated peaks at the station of Zhangjiashan. The results indicate a similar contribution pattern with other scenarios in which the risk indices of  $p_T^{AND}$  and  $p_T^{Kendall}$  are mainly influenced by the parameter uncertainties and also the copula structure. Moreover, the copula structure tends to have a more effect on the predictive variability of  $p_T^{Kendall}$  than that of  $p_T^{AND}$ . For the FP in OR, its variation would be mainly attributed to parameter uncertainty and the marginal structure for the 3-day flood volume (i.e. B). Here the marginal structure (i.e. factor B) for the 3-day flood volume would have a more significant effect on  $p_T^{OR}$  than the marginal structure (i.e. factor A) for the peak flow since the sample data for the 3-day flood volumes would have a higher value of kurtosis (16.6 for 3-day flood volume and 12.0 for flood peak).

In general, the predictive uncertainties for the risk indices of  $p_T^{AND}$  and  $p_T^{Kendall}$  are dominated by the parameter uncertainties and the copula

function, in which the copula structure would have a greater effect on  $p_T^{Kendall}$  than that on  $p_T^{AND}$ . This implies that for flood designs based on  $p_T^{AND}$  or  $p_T^{Kendall}$ , the copula function should be appropriately and then the model parameters need to be well quantified. Table 5 presents the contributions of the copula function to the predictive uncertainties in  $p_T^{AND}$  and  $p_T^{Kendall}$  under a service time of 30-year. It seems that, for the FP in AND, the impact from the copula structure has an opposite trend with the correlation between the correlated variables, but such a phenomenon is not applicable for FP in Kendall. However, this conclusion is only based on limited cases and thus need to be further characterized in future. For the FP in OR, the selection of the marginal distributions would be more important than the copula structure in order to generate reliable risk inferences. Table 6 presents the contributions of the two marginals to the FP in OR under a service time of 30-year. Based on the results, we can primarily conclude that, for the correlated variables, the marginal for the variable with a higher kurtosis would have a greater impact on the predictive uncertainties in  $p_T^{OR}$  than the other marginal.





**Fig. 12.** Contributions of different uncertain factors to the variability of failure probabilities for the 3-day annual maxima volume and the associated flood peak at Zhangjiashan station.

**Table 5**

The copula contributions to failure probabilities in AND and Kendall under a service time of 30-year.

	Kendall's Tau	FP in AND	FP in Kendall
Peak - Volume at Xianyang	0.690	45.9%	53.4%
Peak-Volume at Zhangjiashan	0.723	38.2%	61.7%
3-day Volume-Peak at Zhangjiashan	0.824	23.1%	32.1%

## 6. Conclusions

Extensive uncertainties exist in multivariate risk inferences, which are embedded in marginals, dependence structures and also model parameters. Some studies have been proposed to address parameter uncertainties in the copula-based multivariate risk models. However, the impacts from marginal and copula structures are somewhat overlooked. Therefore, a factorial multimodel Bayesian copula (FMBC) approach has

**Table 6**

The contributions of marginals to failure probabilities in OR under a service time of 30-year.

	Variable	Kurtosis	FP in OR
Peak - Volume at Xianyang	Peak	26	28.9%
	Volume	6.3	18.4%
Peak-Volume at Zhangjiashan	Peak	6.2	5.9%
	Volume	3.1	1.1%
3-day Volume-Peak at Zhangjiashan	Peak	12.0	2.4%
	Volume	16.6	45.6%

been proposed to quantify uncertainties in multivariate risk inferences and further partition the dominant contributors to the uncertain risk inferences. In FMBC, different marginal and copula structures are firstly specified, which lead to different risk inference models. For each model, the AM-based MCMC approach is employed to quantify the parameter uncertainties. The multilevel factorial analysis method is adopted to

characterize the contributions of model marginals, copula structures and the associated parameter uncertainties to the failure probabilities. Moreover, a copula-based dependent sampling algorithm is proposed to jointly sample the parameter values in the factorial analysis process.

The proposed FMBC approach was demonstrated for the multivariate flood risk inferences at two gauging stations in the Wei River basin. Contributions from marginal distributions, dependence structures, as the associated model parameters are characterized to the predictive risks of failure probabilities in AND, OR, and Kendall. Based on those case studies, some specific findings can be concluded:

1. Uncertainties in marginal and dependence structures as well as the associated model parameters would lead to imprecise predictions for the multivariate return periods in AND, OR and Kendall for a compound extreme event.
2. For a specific extreme event, different uncertainties would have different contributions to different risk indices. In general, the copula structure and parameter uncertainties would have the most significant impacts on the failure probabilities in AND and Kendall, while other factors would not have visible contributions (less than 10%) to predictive uncertainties in these two risk indices. Moreover, the contributions of parameter uncertainties will decrease with an increase in the service time while the contributions of the copula will increase. For the failure probability in OR, its uncertainty would be mainly attributed to model parameters and the marginal distributions, but the copula structure will not have a noticeable impact on this risk index. Moreover, when the parameter uncertainties are not well quantified, these uncertainties would dominate the imprecise predictions in the failure probability in OR.
3. For the copula-based multivariate risk model, its parameters in both marginal and copula functions may be correlated with each other. Negligence of their correlation would overestimate the impacts of parameter uncertainties especially for the failure probability in OR. Consequently, the copula-based dependent sampling algorithm is recommended in the factorial analysis process. However, the sample size would not significantly influence the results of uncertainty partition and 10 samples are sufficient to generation reliable characterization for the impacts from different factors.

This study is a first attempt to track the major contributors to the hydrological risk inferences within a multivariate context. The applicability of the FMBC approach has been demonstrated through the flood risk assessment problems under consideration of flood peak and volume. We argue that the proposed method can be applicable for multivariate risk inference problems for many other hydro-climatic extremes with more than two correlated variables.

#### CRedit authorship contribution statement

**Y. Fan:** Study Design, Methodology and Code Development, Calculation, Paper Writing, Funding acquisition. **L. Yu:** Data survey, Funding acquisition. **X. Shi:** Paper proofreading and editing.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgement

This work was jointly funded by the Brunel University Open Access Publishing Fund, and the Royal Society International Exchanges Program (No. IES\R2\202075). We are very grateful for the editor's and the anonymous reviewers' insightful and constructive comments. The data used in this study are collected from the Hydrologic Data Year Book of

the Yellow River and can be accessed at [https://www.researchgate.net/profile/Yurui\\_Fan/publications](https://www.researchgate.net/profile/Yurui_Fan/publications).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2021.126406>.

#### References

- Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44, 182–198.
- Brechmann, E.C., Schepsmeier, U., 2013. Modeling dependence with C- and D-Vine Copulas: The R package CDVine. *J. Stat. Softw.* 52 (3), 1–27.
- Chebana, F., Ouarda, T.B.M., 2011. Multivariate quantiles in hydrological frequency analysis. *Environmetrics* 22 (1), 63–78.
- Cunnane, C. (1989). "Statistical distributions for flood frequency analysis." Operational Hydro. Rep. No. 33, World Meteorological Org. (WMO), Geneva, Switzerland.
- De Michele, C., Salvadori, G., 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* 108 (D2), 4067. <https://doi.org/10.1029/2002JD002534>.
- Du, T., Xiong, L., Xu, C.Y., Gippel, C.J., Guo, S., Liu, P., 2015. Return period and risk analysis of nonstationary low-flow series under climate change. *J. Hydrol.* 527, 234–250.
- Dung, N.V., Merz, B., Bardossy, A., Apel, H., 2015. Handling uncertainty in bivariate quantile estimation – An application to flood hazard analysis in the Mekong Delta. *J. Hydrol.* 527, 704–717.
- Fan, Y.R., Huang, W.W., Li, Y.P., Huang, G.H., Huang, K., 2015. A coupled ensemble filtering and probabilistic collocation approach for uncertainty quantification of hydrological models. *J. Hydrol.* 530, 255–272.
- Fan, Y.R., Huang, G.H., Baetz, B.W., Li, Y.P., Huang, K., 2017. Development of a Copula-based Particle Filter (CopPF) Approach for Hydrologic Data Assimilation under Consideration of Parameter Interdependence. *Water Resour. Res.* 53 (6), 4850–4875.
- Fan, Y.R., Huang, G.H., Zhang, Y., Li, Y.P., 2018. Uncertainty quantification for multivariate eco-hydrological risk in the Xiangxi River within the Three Gorges Reservoir Area in China. *Engineering* 4 (5), 617–626.
- Fan, Y., Huang, K., Huang, G., Li, Y., 2020a. A factorial Bayesian copula framework for partitioning uncertainties in multivariate risk inference. *Environ. Res.* 183, 109215.
- Fan, Y., Huang, K., Huang, G., Li, Y., Wang, F., 2020b. An uncertainty partition approach for inferring interactive hydrologic risks. *Hydrol. Earth Syst. Sci.* 24 (9), 4601–4624.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiérmonge, N., Bobée, B., 2004. Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* 40, W01101. <https://doi.org/10.1029/2003WR002456>.
- Ganguli, P., Reddy, M.J., 2013. Probabilistic assessment of flood risks using trivariate copulas. *Theor. Appl. Climatol.* 111 (1–2), 341–360.
- Graler, B., van den Berg, M.J., Vandenbergh, S., Petroselli, A., Grimaldi, S., De Baets, B., Verhoest, N.E.C., 2013. Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation. *Hydrol. Earth Syst. Sci.* 17, 1281–1296.
- Guo, A.J., Chang, J.X., Wang, Y.M., Huang, Q., Li, Y.Y., 2020. Uncertainty quantification and propagation in bivariate design flood estimation using a Bayesian information-theoretic approach. *J. Hydrol.* 584, 124677.
- Haario H., Saksman E., Tamminen J., An adaptive Metropolis algorithm. *Bernoulli*, 7(2), 223–242.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application. *Biometrika* 57, 97–109.
- Huang, K., Dai, L.M., Yao, M., Fan, Y.R., Kong, X.M., 2017. Modelling dependence between traffic noise and traffic flow through an entropy-copula method. *J. Environ. Inf.* 29 (2), 134–151. <https://doi.org/10.3808/jei.201500302>.
- Kao, S.C., Govindaraju, R.S., 2010. A copula-based joint deficit index for droughts. *J. Hydrol.* 380, 121–134.
- Karmakar, S., Simonovic, S.P., 2009. Bivariate flood frequency analysis. Part 2: a copula-based approach with mixed marginal distributions. *J. Flood Risk Manage.* 2, 32–44.
- Kidson, R., Richards, K.S., 2005. Flood frequency analysis: assumption and alternatives. *Prog. Phys. Geogr.* 29 (3), 392–410.
- Lee, T., Salas, J.D., 2011. Copula-based stochastic simulation of hydrological data applied to Nile River flows. *Hydrol. Res.* 42 (4), 318–330.
- Li Z., J.J., Li, X.P., Shi, (2020). A Two-Stage Multisite and Multivariate Weather Generator. *Journal of Environmental Informatics*, 35(2), 148–159.
- Liu, Y.R., Li, Y.P., Ma, Y., Jia, Q.M., Su, Y.Y., 2020. Development of a Bayesian-copula-based frequency analysis method for hydrological risk assessment – The Naryn River in Central Asia. *J. Hydrol.* 580, 124349.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state by fast computing machines. *J. Chem. Phys.* 21, 1087–1092.
- Ministry of Water Resources of the People's Republic of China (2006). Regulation for Calculating Design Flood of Water Resources and Hydropower Engineering; China Waterpub Press: Beijing, China, 2006; pp. 1–45. (In Chinese).
- Moeini, R., Soltani-nezhad, M., 2020. Extension of the constrained gravitational search algorithm for solving multi-reservoir operation optimization problem. *J. Environ. Inf.* 36 (2), 70–81.
- Moftakhari, H.R., Salvadori, G., AghaKouchak, A., Sanders, B.F., Matthew, R.A., 2017. Compounding effects of sea level rise and fluvial flooding. *Proc. Natl. Acad. Sci.* 114 (37), 9785–9790.

- Nelsen, R.B., 2006. *An Introduction to Copulas*. Springer, New York.
- Qi, W., Zhang, C., Fu, G., Zhou, H., 2016. Imprecise probabilistic estimation of design floods with epistemic uncertainties. *Water Resour. Res.* 52, 4823–4844. <https://doi.org/10.1002/2015WR017663>.
- Read, L.K., Vogel, R.M., 2015. Reliability, return periods, and risk under nonstationarity. *Water Resour. Res.* 51, 6381–6398. <https://doi.org/10.1002/2015WR017089>.
- Requena, A., Mediero, L., Garrote, L., 2013. A bivariate return period based on copulas for hydrologic dam design: Accounting for reservoir routing in risk estimation. *Hydrol. Earth Syst. Sci.* 17 (8), 3023–3038.
- Sadegh, M., Moftakhari, H., Gupta, H.V., Ragno, E., Mazdiyasi, O., Sanders, B., Matthew, R., AghaKouchak, A., 2018. Multihazard scenarios for analysis of compound extreme events. *Geophys. Res. Lett.* 45, 5470–5480.
- Salvadori, G., De Michele, C., Durante, F., 2011. On the return period and design in a multivariate framework. *Hydrol. Earth Syst. Sci.* 15, 3293–3305.
- Salvadori, G., Durante, F., De Michele, C., 2013. Multivariate return period calculation via survival functions. *Water Resour. Res.* 49, 2308–2311. <https://doi.org/10.1002/wrcr.20204>.
- Salvadori, G., De Michele, C., Kottegoda, N.T., Rosso, R., 2007. *Extremes in Nature: an Approach using Copula*. Springer, Dordrecht, p. 292.
- Salvadori, G., Durante, F., De Michele, C., Bernardi, M., Petrella, L., 2016. A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities. *Water Resour. Res.* 52, 3701–3721. <https://doi.org/10.1002/2015WR017225>.
- Sarhadi, A., Burn, D.H., Ausin, M.C., Wiper, M.P., 2016. Time-varying nonstationary multivariate risk analysis using a dynamic Bayesian copula. *Water Resour. Res.* 52, 2327–2349. <https://doi.org/10.1002/2015WR01852>.
- Serinaldi, F., 2013. An uncertain journey around the tails of multivariate hydrological distributions. *Water Resour. Res.* 49, 6527–6547. <https://doi.org/10.1002/wrcr.2053>.
- Song, J., Xu, Z., Liu, C., Li, H., 2007. Ecological and environmental instream flow requirements for the Wei River – the largest tributary of the Yellow River. *Hydrol. Process.* 21, 1066–1073.
- Sraj, M., Bezak, N., Brilly, M., 2014. Bivariate flood frequency analysis using the copula function: a case study of the Litija station on the Sava River. *Hydrol. Process.* 29 (2), 225–238. <https://doi.org/10.1002/hyp.10145>.
- Sun C.X., G.H., Huang, Y.R., Fan, X., Zhou, C., Lu, X.Q., Wang, (2019). Drought occurring with hot extremes: Changes under future climate change on Loess Plateau, China. *Earth's Future* 7 (6), 587–604.
- Sun C.X., G.H., Huang, Y.R., Fan, X., Zhou, C., Lu, X.Q., Wang, (2021). Vine copula ensemble downscaling for precipitation projection over the Loess Plateau based on high-resolution multi-RCM outputs. *Water Resources Research*, 57(1), e2020WR027698, <https://doi.org/10.1029/2020WR027698>.
- ter Braak, C.J.F., Vrugt, J.A., 2008. Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing* 18, 435–446.
- The European Parliament and The Council (2007), Directive 2007/60/EC: On the assessment and management of flood risks, Official Journal of the European Union, 116 pp.
- Viglione, A., R. Merz, J.L. Salinas, G. Blischl (2013). Flood frequency hydrology: 3. A Bayesian analysis, *Water Resources Research*, 49, 675–692.
- Vogel, R., Wilson, I., 1996. The probability distribution of annual maximum, minimum and average streamflow in the United States. *J. Hydrol. Eng.* 1, 69–76. [https://doi.org/10.1061/\(ASCE\)1084-0699\(1996\)1:2\(69\)](https://doi.org/10.1061/(ASCE)1084-0699(1996)1:2(69)).
- Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdón, D., 2009. Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation* 10, 273–290.
- Xu, Y., Huang, G.H., Fan, Y.R., 2016. Multivariate flood risk analysis for Wei River. *Stoch. Env. Res. Risk Assess.* 31 (1), 225–242. <https://doi.org/10.1007/s00477-015-1196-0>.
- Yue, S., 2000. The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol. Process.* 14 (14), 2575–2588.
- Yue, S., 2001. A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol. Process.* 15 (6), 1033–1045.
- X. Zhou G.H., Huang, X., Wang, Y.R., Fan, G.H., Cheng, A coupled dynamical-copula downscaling approach for temperature projections over the Canadian Prairies *Climate Dynamics* 51 7 2018 2413 2431.
- X.W. Zhuang Y.P., Li, S., Nie, Y.R., Fan, G.H., Huang, Analyzing climate change impacts on water resources under uncertainty using an integrated simulation-optimization approach *Journal of Hydrology* 556 2018 523 538.
- Zscheischler, J., Seneviratne, S.I., 2017. Dependence of drivers affects risks associated with compound events. *Sci. Adv.* 3, e1700263.
- Zuo, D.P., Xu, Z.X., Wu, W., Zhao, J., Zhao, F.F., 2014. Identification of streamflow response to climate change and human activities in the Wei River basin, China. *Water Resour. Manag.* 28 (3), 833–851.